Jet Propulsion Laboratory
California Institute of
Technology

Data and Model Analysis and Uncertainty Quantification for Sea Level Science

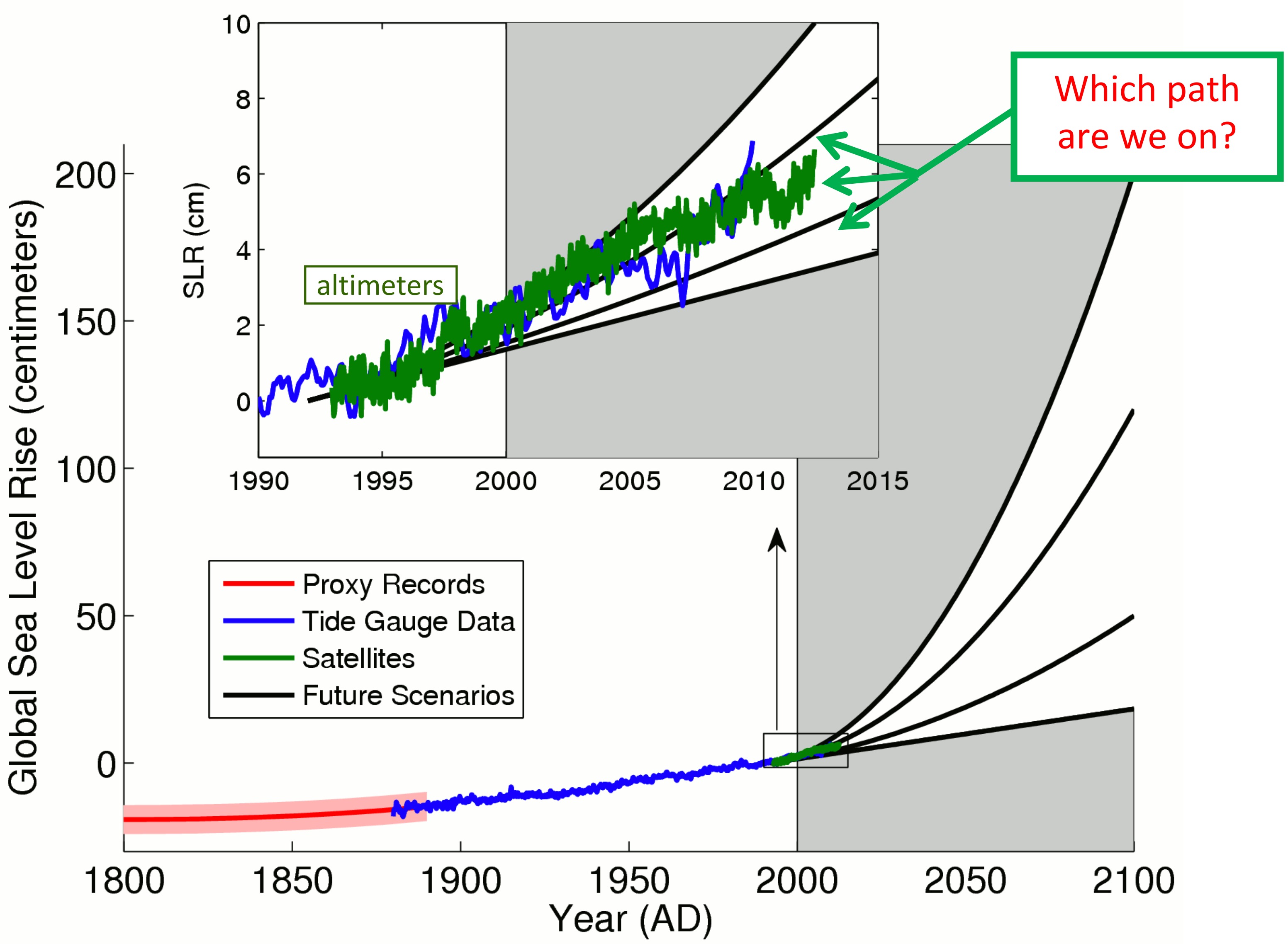
Carmen Boening

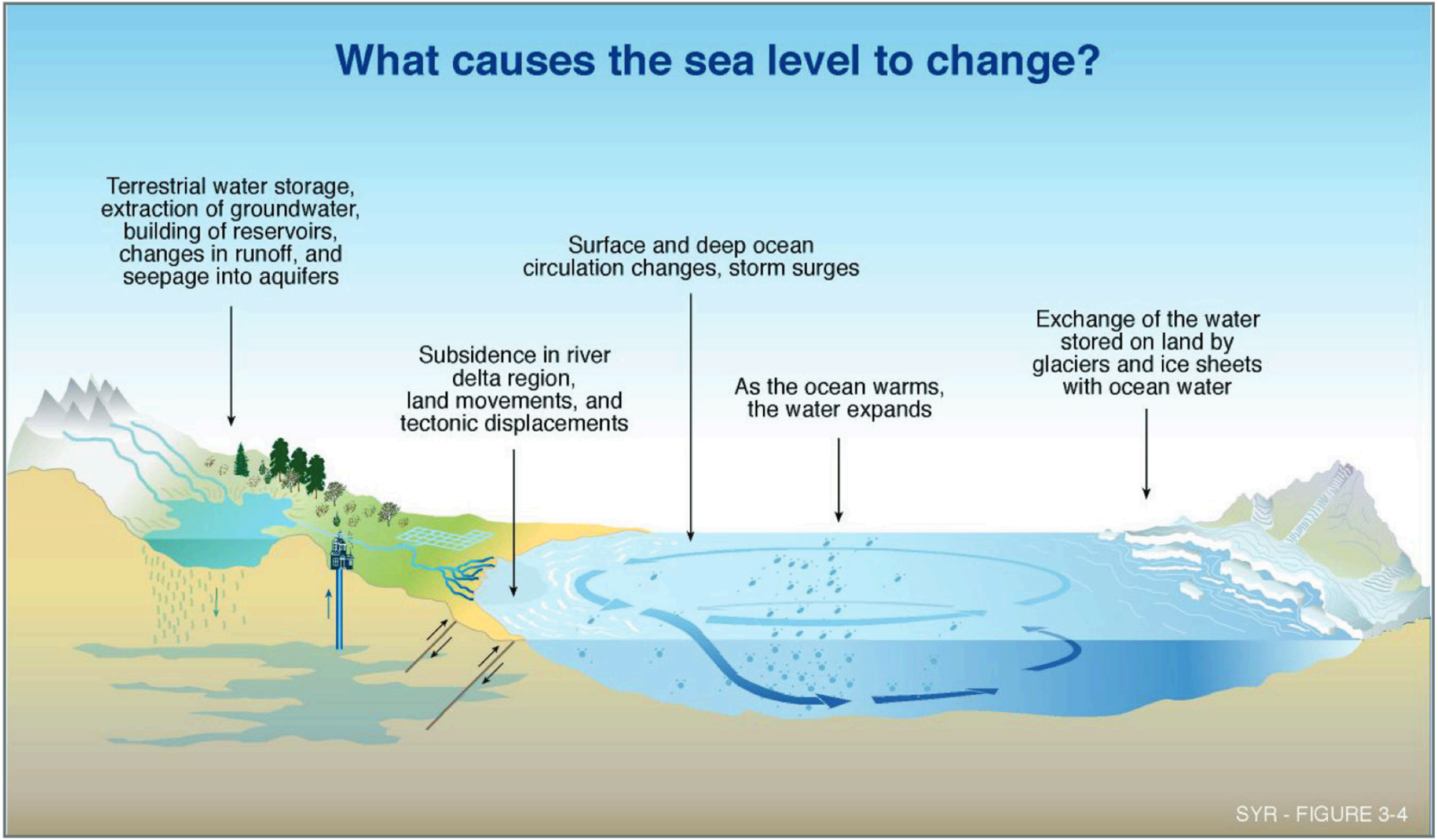
Jet Propulsion Laboratory, California Institute of Technology

Pasadena, CA

Sea Level Rise Projections

Jet Propulsion Laboratory
California Institute of Technology




What causes the sea level to change?
Terrestrial water storage, extraction of groundwater, building of reservoirs, changes in runoff, and seepage into aquifers
Surface and deep ocean circulation changes, storm surges
Subsidence in river delta region, land movements, and tectonic displacements
As the ocean warms, the water expands
Exchange of the water stored on land by glaciers and ice sheets with ocean water
SYR - FIGURE 3-4

Common challenges

Analysis requires combining data from:

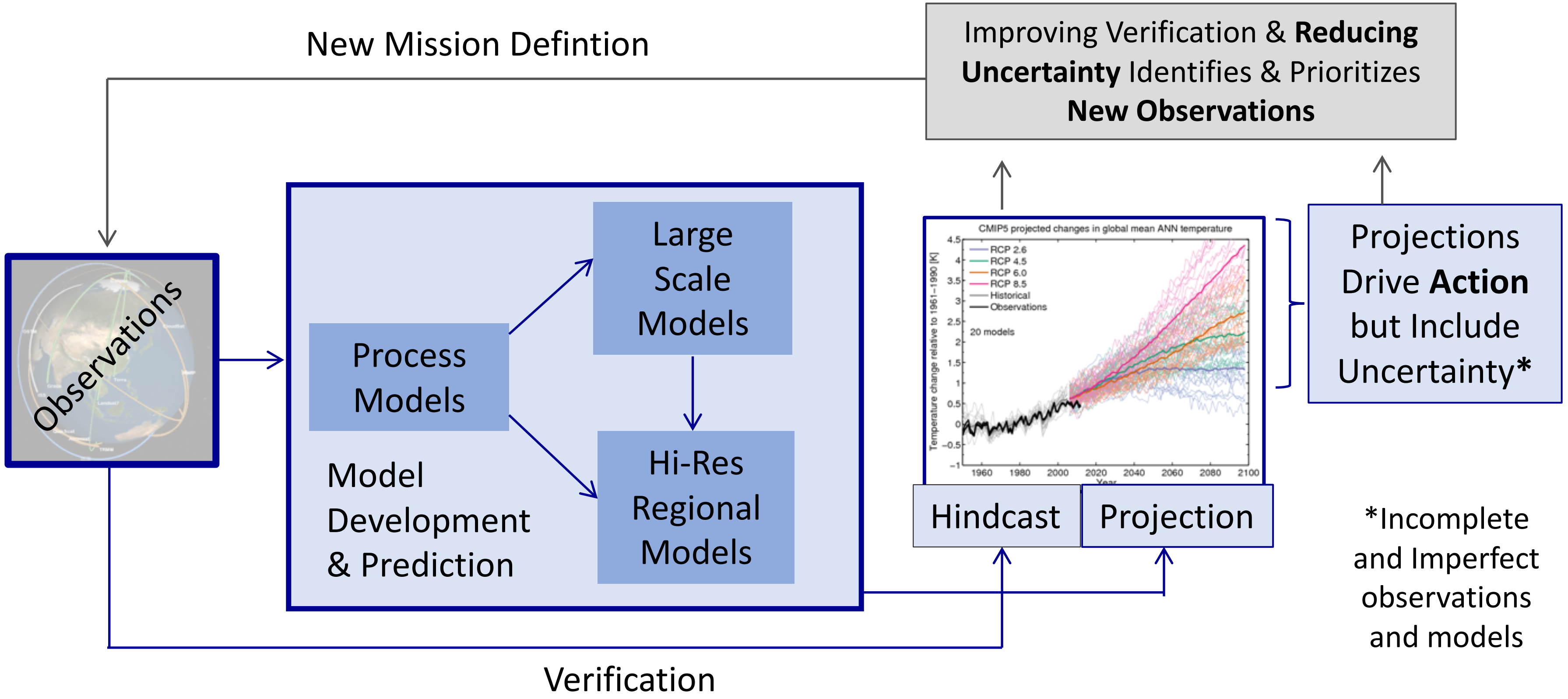
- Large variety of data sources
- Different formats
- Large datasets
- Datasets that vary in latency
- Datasets with a variety of sources of uncertainties

Gridding and interpolation in data combination requires sophisticated statistical methods to compute uncertainties

Jet Propulsion Laboratory
California Institute of Technology

Science System Engineering:

Providing high accuracy answers, with quantitative uncertainty and end-to-end traceability



*Incomplete and Imperfect observations and models

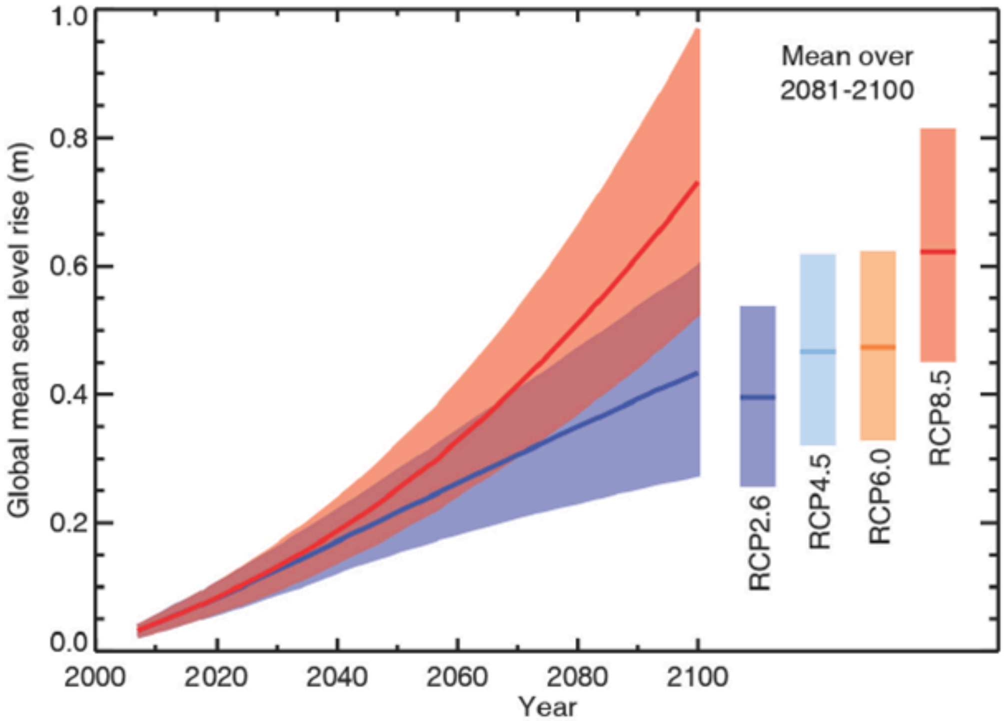
- Better quantitative characterization of these complex systems through the application of system engineering and uncertainty quantification methods would enable:
  - Improved *science analysis* results
  - Improved *science traceabiity* for optimizing measurement system (mission and instruments) design
  - Improved *prioritization* of missions and instruments

Over-all Modelling Experiment Approach

- Improve SLR prediction via refinement / investigation of ice sheet contributions using JPL ISSM (ice sheet model) & MITGCM (ocean model)
  - Atmospheric (SMB) and ocean (melt rate) effects on the ice sheet are rolled in via mixture of sensitivity studies and climate scenario runs
  - Investigate credibility of high end melt rate scenarios by doing ocean / ice interface modelling & sensitivity experiments using MITgcm / ECCO 2
- Focus on worst case / end member scenarios
  - When faced with very large uncertainty in a design problem, a typical engineering response is to pick known worst case values, and/or apply substantial margin (Model Uncertainty Factor – MUF)
    - There is a large amount of uncertainty on what the atmospheric and ocean contributions to ice sheet evolution -> both in credible warming scenarios, and impacts of those scenarios
  - Prioritized follow up studies for providing additional insight into upper bounds
    - In places where the conservative worst case value has large impact (i.e. predict is too large to be “helpful”, per previous slide), investigate if more credible values can be selected based on sound physical principles and/or sensivity studies of driving phenomena
    - Investigate / quantify additional potential first order feedback phenomenon
      - E.g. GIS Ice sheet shrinking resulting in lower altitude ice sheet and hence more rain less snow…
      - Still thinking about this

Jet Propulsion Laboratory
California Institute of Technology

An Example: Projecting Sea Level Rise: A Grand Challenge Science Question



Projecting Sea Level Rise is a critical and Grand Challenge Earth **System** Science Question

JPL owns the global satellite measurements for:

- sea level & its rise (e.g. Jason)
- distinguishing between mass input - melting ice & snow, and thermal expansion (i.e. Jason & GRACE)

JPL is home to leadership in all science areas critical to sea level rise:

- oceanography, natural climate variability, cryosphere and ice sheets, solid earth and post glacial rebound, gravity variations

Addressing this question, among others in Earth and Space Science, requires developing an **advanced and combined science and system engineering capability** to *quantitatively* trace existing and proposed scientific measurement to key science questions.

Jet Propulsion Laboratory
California Institute of Technology

SLR-P Project Objectives

1. **Produce a [95% confidence] GMSLR upper bound for the year 2100**
2. **Generate prioritized list of measurements that will help reduce uncertainties in GMSLR projections**
   - Priorities based on *quantified* assessment of highest impact on prediction uncertainty reduction
3. **Start building coalition / work stakeholder development to:**
   - look for long term funding sources outside of RTD
   - Help begin the process of changing the culture of how "science with societal impacts" problems are worked -> transition to structure more in line with flight program / project approach -> more dedicated resources, SE to help with coordination and priorities, and management; less volunteer and part time
   - Not substantially started at this time

Jet Propulsion Laboratory
California Institute of Technology

What Is The Current State of The Practice?

- **SLR uncertainty estimates** state of the practice
  - Subject matter expert polling approach (e.g. IPCC, NCA) – leads to mixing numbers from a variety of disparate approaches
  - Pfeffer et al "kinematic constraints" approach – based on what we know of glacier dynamics, what is the most mass one could credibly imagine could come out of ice-sheets ("worst-case")
  - Some sensitivity experiments and model intercomparisons have been done
    - SeaRISE – 2009 to ~2014
      - Initial ice sheet model intercomparison project (voluntary basis – no funding attached)
    - ISMIP, MISMIP, MISOMIP… *in work* for CMIP-6
      - Requirements for input/output parameters and boundary conditions
    - Looking at bounding errors and uncertainties, partially by comparing different models to each other (mainly looking at model correlations)
    - *Ideally* you give each model the exact same inputs and boundary conditions, and compare outputs.
    - *Challenging* to do since different ice sheet models have different formulations, so inputs and BCs are slightly or substantially different, so interpretation of resultant output differences can be challenging – lack of coordination partly due to funding situations of these efforts
  - To our knowledge, no one has done continent scale UQ run with full statistics, coupling in meaningful uncertainties (scaled by sensitivity) on key parameters / BCs
- **Quantitative Measurement priorities** – never done for ice sheet measurements to our knowledge
  - However, many, over-lapping, papers / books on "batch" of important measurements to make based on expert knowledge
  - No quantitative justification, and some community reluctance to create priorities
- Literature review and WCRP / national climate community group plans indicate that the community is slowly becoming more rigorous and quantitative – decade+ long trend
  - Opportune time for JPL to help lead / push / accelerate this trend

SLR Upper Bound – Approach & UQ Challenge

- Understand strengths and weaknesses of models being used (PIRT)
  - Long term help ID priorities for model dev / adding missing physics / measurements
- Generate experiment matrix based on a-priori estimate of ice sheet parameters / BCs with biggest impact on SLR
  - Ideally would run experiments on all, but not computationally feasible within scope of RTD
- Conduct sensitivity experiments to identify most important parameters / BCs impacting SLR
  - Use "extreme worst case" values to see if a useful upper bound constraint on SLR can be derived
    - Done, answer is mostly NO – stacked worst case results in 1.3 to 2.5m SLR from AIS and GIS, vs ~1.2m from Pfeffer et al high end (our reference target, since this was the source for worst case value in the National Climate Assessment)
  - Re-do with more credible worst case values – in process
- Based on sensivity experiment results, do full scale UQ run for AIS and GIS
  - Investigate effect of partitions
  - Use conservative / easy-to-defend parameter value distributions
- Key challenge is how to incorporate what we know of uncertainties for parameters and BCs, and how they vary regionally, into models / UQ math in a physically meaningful way
  - Would like to discuss this with Peterson, Braverman, ISSM modelling team

Jet Propulsion Laboratory
California Institute of Technology

Why and where does UQ fit in to advance the field or realize the project's objective?

- UQ provides vital framework for both key objectives of the SLR-P RTD
  - Used for bounding SLR predictions for 2100 and reducing uncertainties in current predictions
  - Used for identifying model development priorities
    - What are the the missing physics that impact the prediction the most?
  - Used for identifying measurement priorities
    - Which parameter is not well defined and has a big impact on the uncertainty in the prediction?
    - Identify measurement requirements (e.g. temporal and spatial resolution)
- Framework can be used to make requirements/uncertainties more traceable and make decisions less “opinion-based”
  - Implementation of a paradigm change in the field (potentially bigger, more focused projects with full-time FTEs)

Successes and Challenges – SLR Projection

- Key successes / insights
  - Conducting extreme worst case experiments can be very insightful – for example
    - Greenland glaciers and the ocean i/f – relatively small total sensitivity of SLR to even massive melting of GIS outlet glaciers (~15cm SLR in 100 yrs)
    - Even with stacked worst case estimates, if it very difficult to see how SLR by 2100 would exceed 3-4m -> Hansen et al claims
  - Errors / uncertainties in AIS bed topography maps have potentially very significant impacts on ability to predict future (e.g. 30% different SLR predict for stacked worst case conditions depending on which bed topo model is used)
    - No major call for icesheet measurements has highlighted need for better bed topography knowledge -> JPL recently pushed for this in recent DS white paper
- Biggest challenges
  - Ocean sensitivity experiments are computationally intense and have substantial run unique startup "NRE" -> about 75% done with cavity experiments after ~1+ yr of work
  - Being smart(er) about credible worst cases is much more difficult
    - Getting to defendable 95% upper bound on SLR by 2100 which is better then the Pfeffer et al 2m estimate might not be possible within the scope of work we have remaining in RTD
    - Really need more people / money / time to do these things right - collaboration with others
  - Statistical tools exist to facilitate UQ – e.g. DAKOTA – but these don't solve the garbage in / garbage out issue with models and uncertainty estimates – see next page
    - Getting "appropriate" or "physically real" uncertainty estimates requires the next level of mapping uncertainties and statistical variation. Need to:
      - Create icesheet partitions in ISSM that map to real climate regions, statistically vary parameters / BCs in these partition areas to get higher quality UQ estimates
      - map climate scenario ranges to the partitions, map melt rate ranges to the partitions

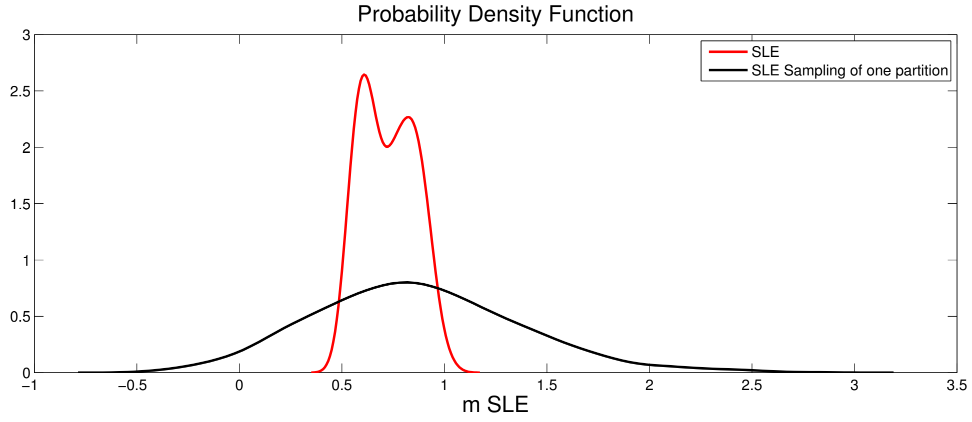
    - Need to work with ISSM team, Braverman, Peterson to see how to apply our knowledge and improve

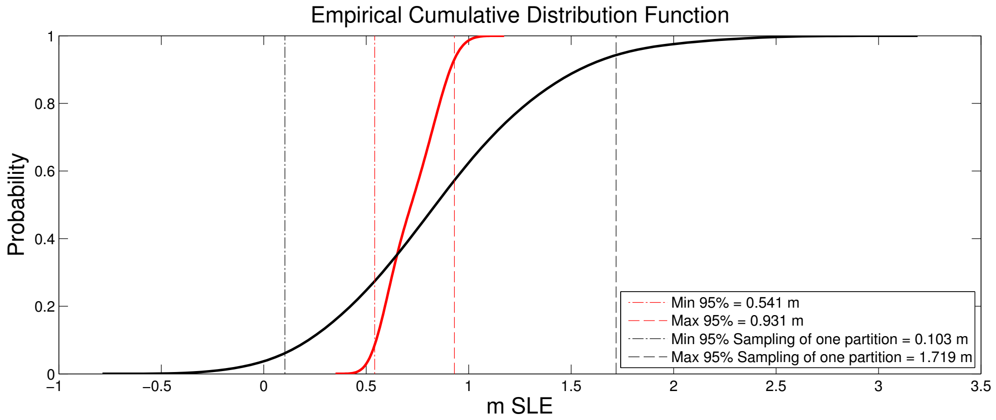
Jet Propulsion Laboratory
California Institute of Technology

Example of impact on partition choice on statistics

CASES:						
Parameter	**Min**	**Max**	**Distribution type**	**Sampling**	**# of cases**	**Notes**
AIS:						
Ocean I/F Melt	"min melt" from Schodlok	control melt * 10	uniform	Latin hypercube	200	ok for now, conservative
Basal Drag	**control case values**	-60%	uniform	Latin hypercube	200	ok for now, conservative
Viscosity	control case values	-40%	uniform	Latin hypercube	200	ok for now, conservative
Accumulation	-50%	+100%	uniform	Latin hypercube	200	ok for now, conservative
Run-off	0	0	NA			
Total cases					800	
CPU hrs per case					45	200 yr 2D
CPU hrs per case					22.5	100 yr 2D
Total CPU hrs					36000	200 yr
Total CPU hrs					18000	100 yr 5%

- Sampling of physically possible but not necessarily realistic worst-cases provides conservative upper bound (TBC)
- *Next*:
    - UQ runs for realistic future scenarios based on climate projections
    - Smarter partitions





Jet Propulsion Laboratory
California Institute of Technology

Successes and Challenges – Measurement Priorities

- Key successes / insights
  - SLR response sensitivity to bed topography errors is relatively straightforward to set up and run -> quantifiable benefit
    - Starting to work with Ice Penetrating Radar team to help them justify science merit of measurement

- Biggest challenges
  - Remaining space based measurements are difficult to provide quantified values for – accuracy, resolution, frequency
    - Velocity (InSAR), Topography (radar, laser)
    - Quantifiable impacts based on current models would show moderate to negligible benefit based on current physics
    - Main argument for these measurements is to deepen process understanding and ID modelling of currently missing physics, so almost by definition current models cant be used to quantify the benefit of missing process understanding / physics
    - Looking into “analog” approach – e.g. “if we uncover missing physics with an impact similar to the last major model upgrade that was made, then SLR uncertainty / error might be reduced by XX”
  - Other substantial benefit of new and existing measurements is data assimilation – Have not discussed a framework for this (Larour might have one?) and it could be used to help defend measurement priorities, help validate models…
- One potential approach for making measurement recommendations is risk based:
  - “IPR is a known / high payoff measurement; others MIGHT be very useful or might not be. Do IPR first, then the others…”

Jet Propulsion Laboratory
California Institute of Technology

Traditional Method for Analyzing Satellite Measurements

Search → Download → Compute

Depending on the data volume (size and number of files)

It could take many hours of download – (e.g. 10yr of observational data could yield thousands of files)

It could take many hours of computation

It could require expensive local computing resources

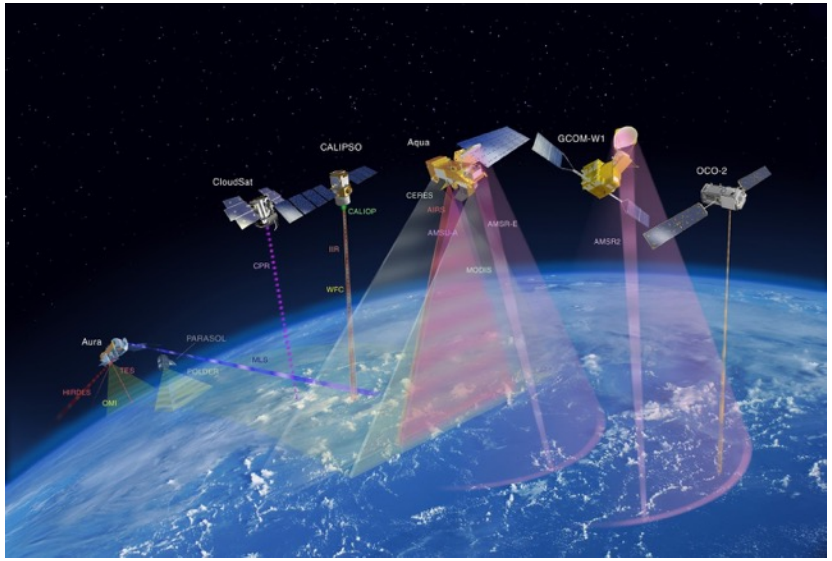
Often after result is produced, purge downloaded files

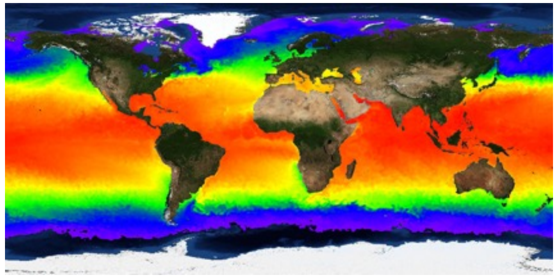
Observations

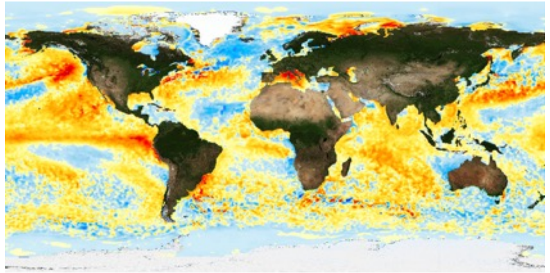
Traditional methods for data analysis (time-series, distribution, climatology generation) can't scale to handle large volume, high-resolution data. They perform poorly

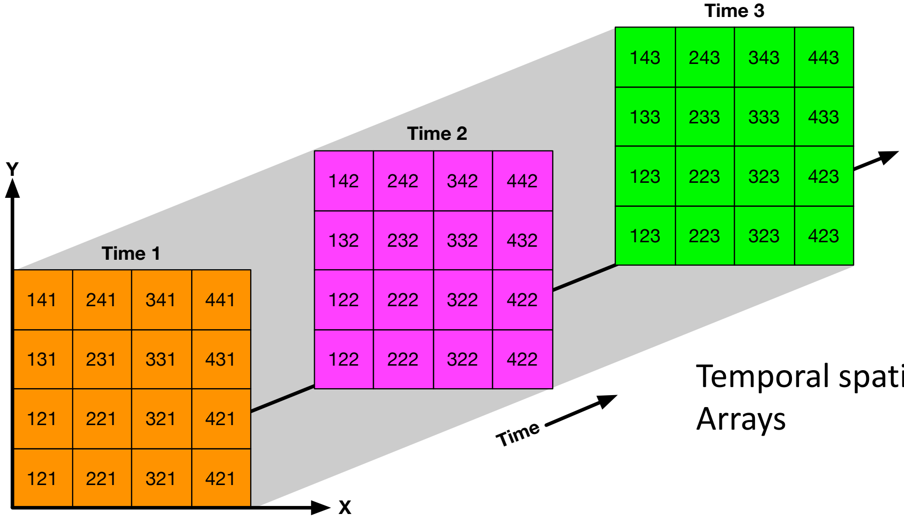
Performance suffers when involve large files and/or large collection of files

A high-performance data analysis solution must be free from file I/O bottleneck









Temporal spatial Arrays

NASA Sea Level Change Portal – https://sealevel.nasa.gov

Analyze contributions to past, current and future regional Sea Level Change

- Determine how much will sea level rise by [2100]?
- What are the key sensitivities?
- Where are the key uncertainties?
- Where are the key Observables? Model Improvements

Goals for the NASA Sea Level Change Portal

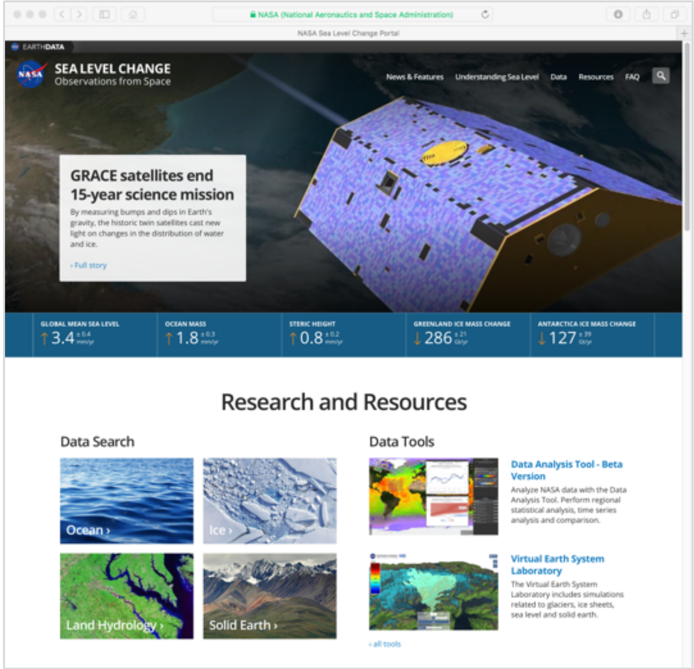
- Provide scientists and the public with a “one-stop” source
- Provide current sea level change information and data
- Provide interactive tools for analyzing and viewing regional data
- Provide virtual dashboard for sea level indicators
- Provide latest news, quarterly report, and publications
- Provide ongoing updates through a suite of editorial products

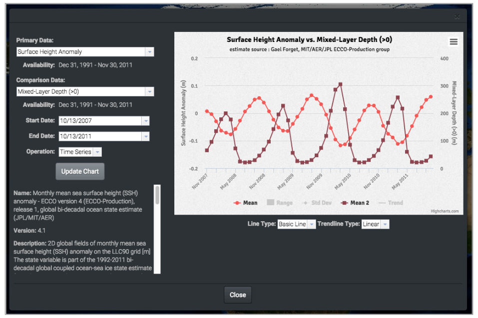
Requires

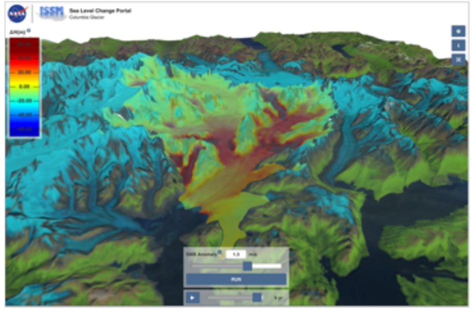
- Interdisciplinary collaboration
- Connect disciplines and evaluate dependencies

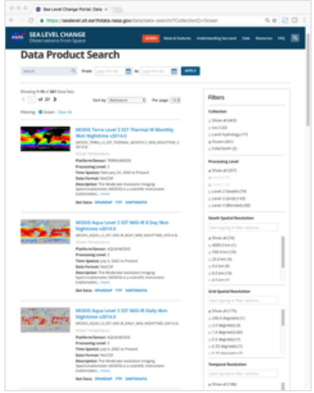
Sea Level Change Portal facilitates

- Easy interdisciplinary data comparison
- Access to latest news and information
- Collaboration (data and information exchange)





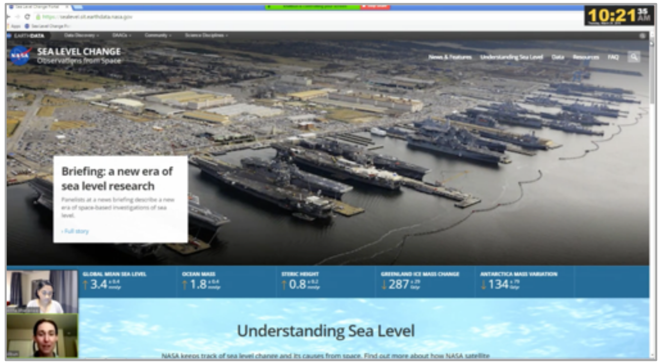


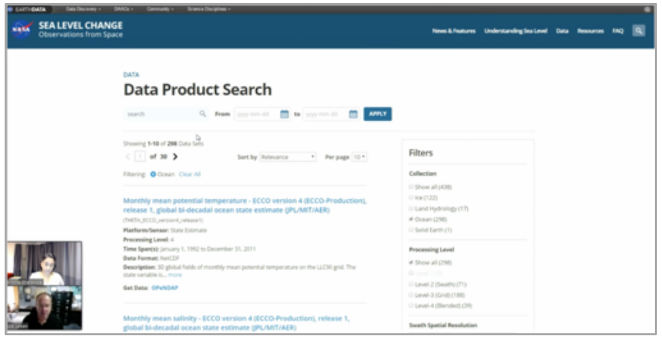


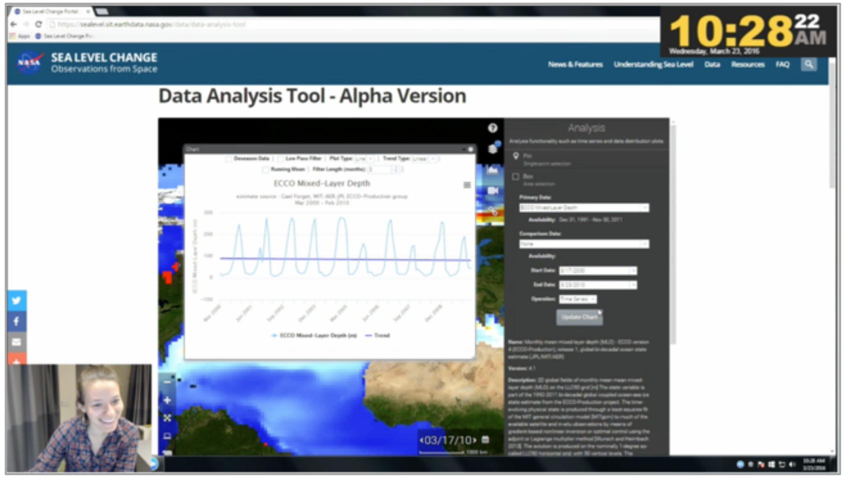
Jet Propulsion Laboratory
California Institute of Technology

UX Assessment

- **Understand** how and for what purposes users obtain sea level change data and information
- **Describe** users' pain points and unmet needs for extracting, visualizing, comparing and analyzing sea level change data
- **Investigate** users' current practices, interests and preferences for engaging in interaction with other sciences and the public about sea level change
- **Identify** opportunities for enhancing cross-disciplinary collaborative activities on the web portal.

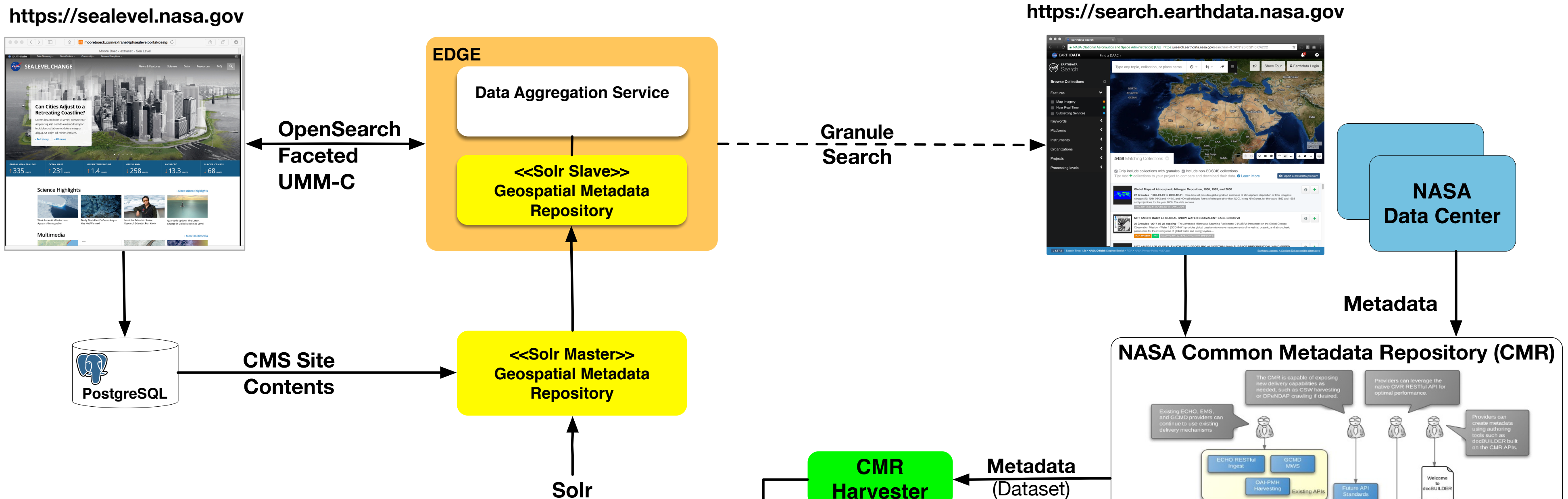


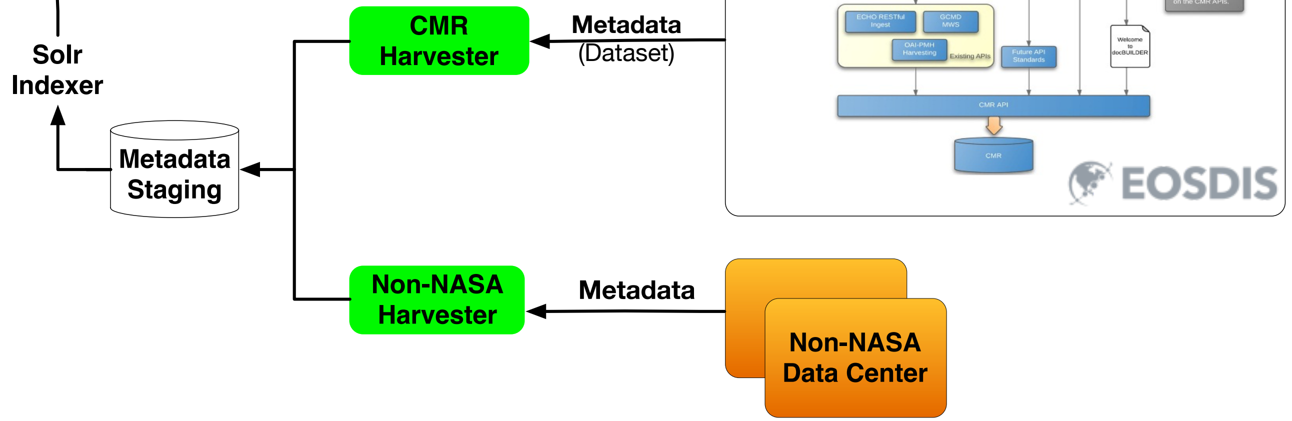




Jet Propulsion Laboratory
California Institute of Technology

One-Stop Search





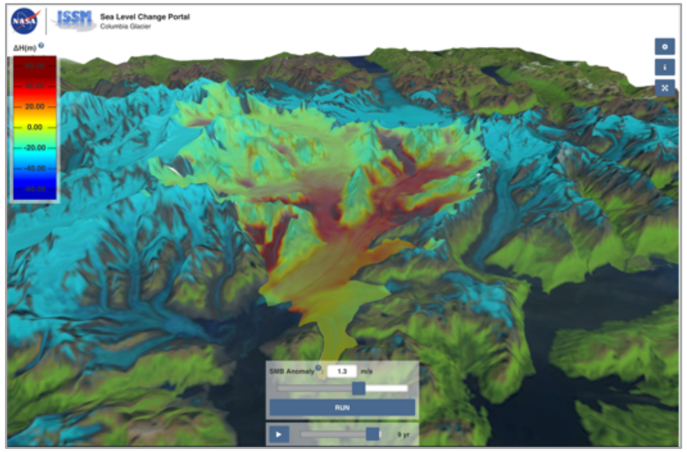
- Build on Free and Open Source Solutions
- Master-slave architecture to prevent data corruption
- Extensible architecture to support various metadata standards
- Embrace OpenSearch, Faceted browse, and Unified Metadata Model for Collections (UMM-C)

Virtual Earth System Laboratory (VESL)

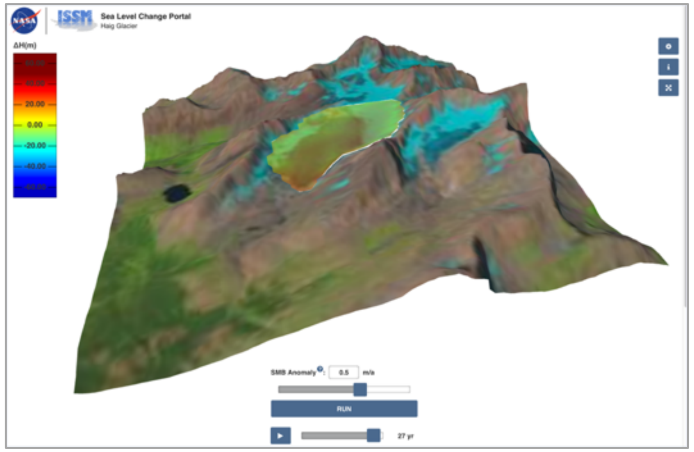
Jet Propulsion Laboratory
California Institute of Technology

Suite of Interactive Simulation Tools for Glaciers, Ice Sheets, Sea Level, and Solid Earth

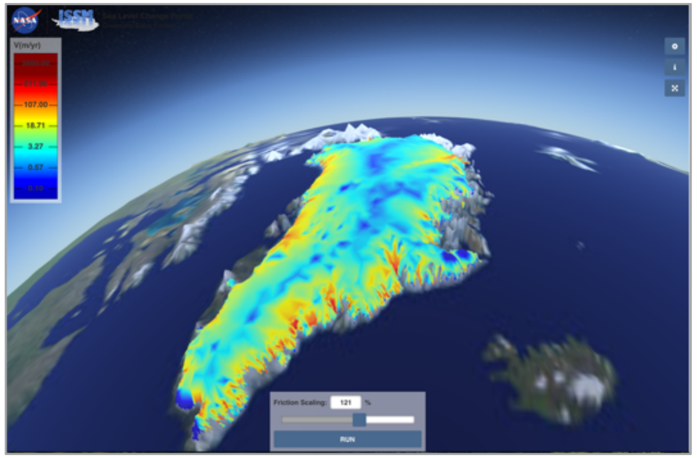
Simulation Service on Amazon Cloud



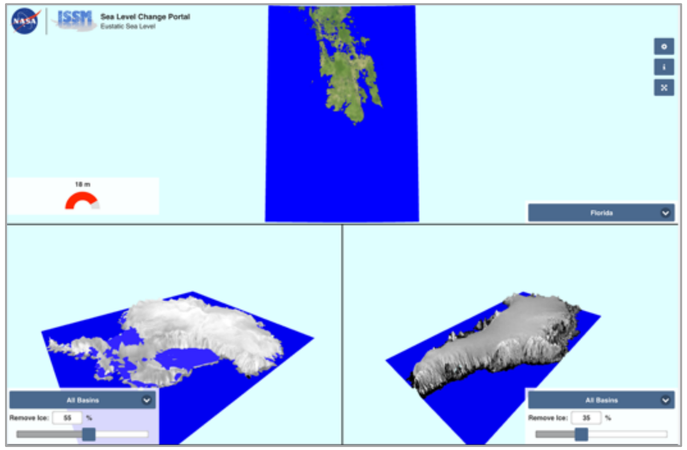
Columbia Glacier



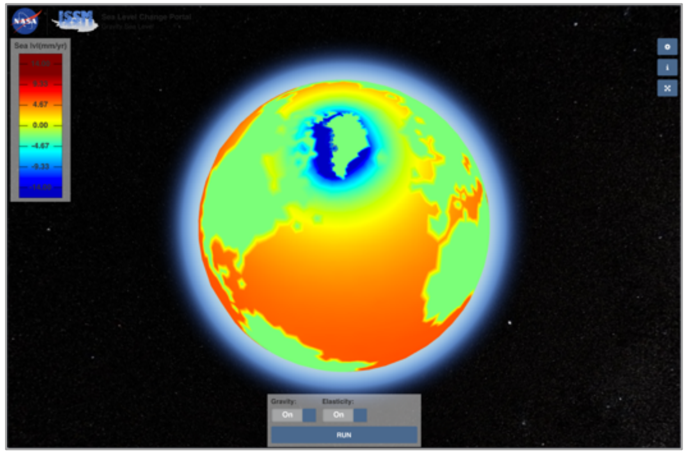
Haig Glacier



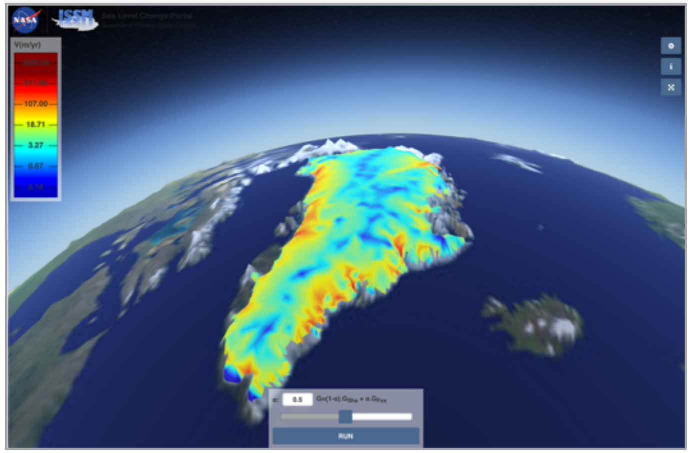
Greenland Basal Friction



Eustatic Sea Level



Local Sea Level

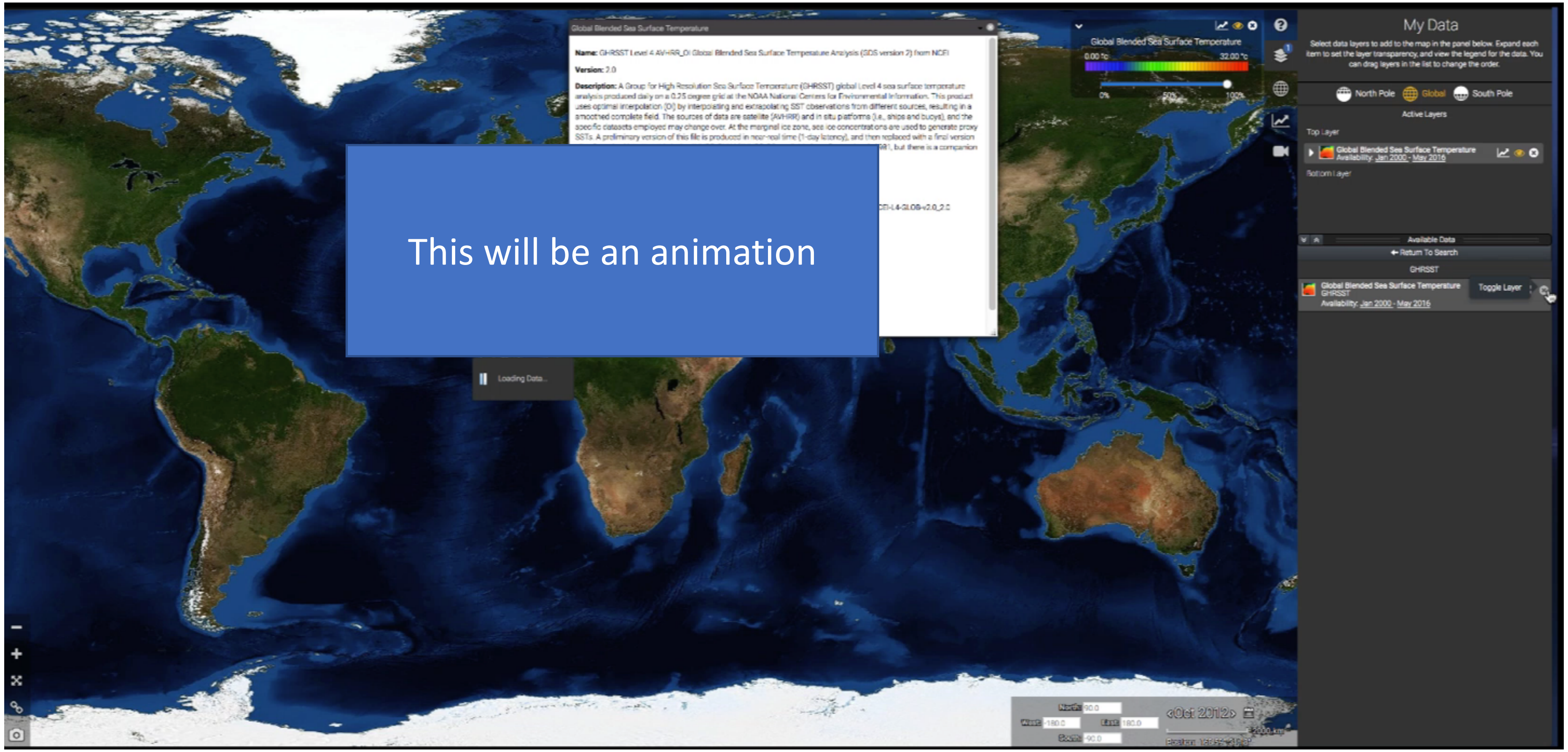


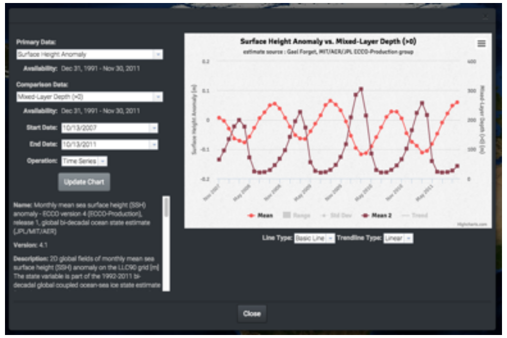
Greenland Steady-State-Friction

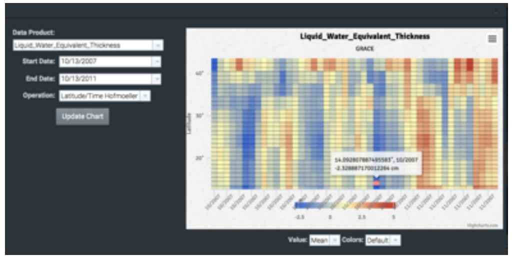
NASA Sea Level Change Portal

Jet Propulsion Laboratory
California Institute of Technology

Analyze Sea Level On-The-Fly *https://sealevel.nasa.gov*







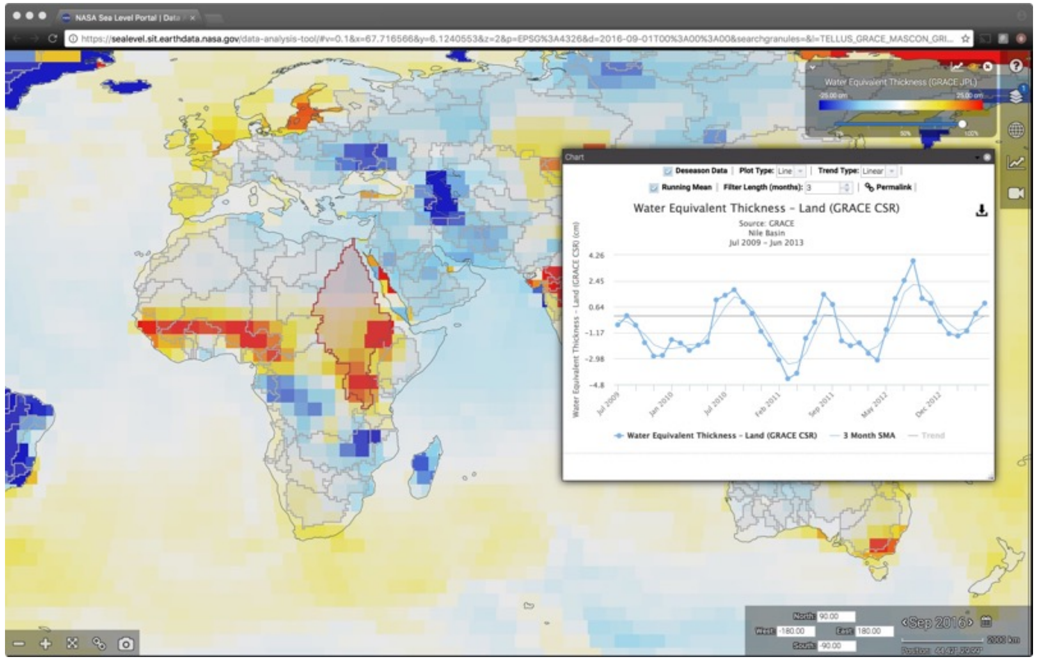
Sea Level Change - Data Analysis Tool

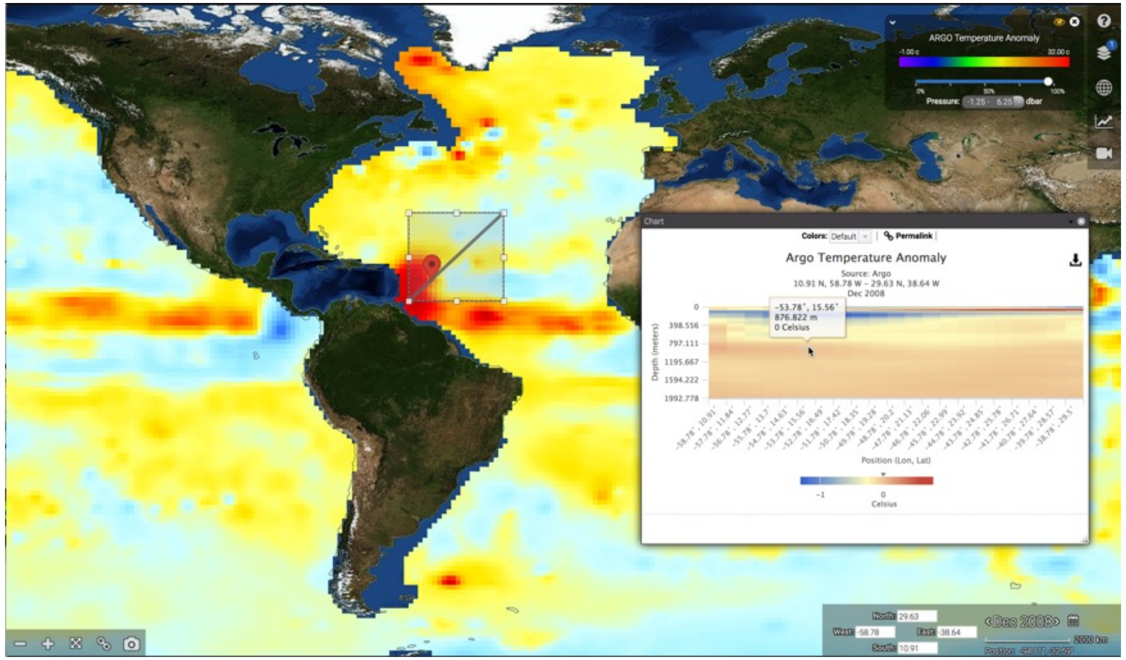
Visualizations | Animations | Polar Projection | Hydrological Basins | In Situ | Time Series | Deseason | Data Comparison | Scatter Plot | Latitude/Time Hovmöller | Etc.

Hydrological Basin and ARGO Analysis

- Browse metadata on hydrological basins
- Enable analysis on hydrological basins
- Supports on-the-fly deseasoning and running mean computation

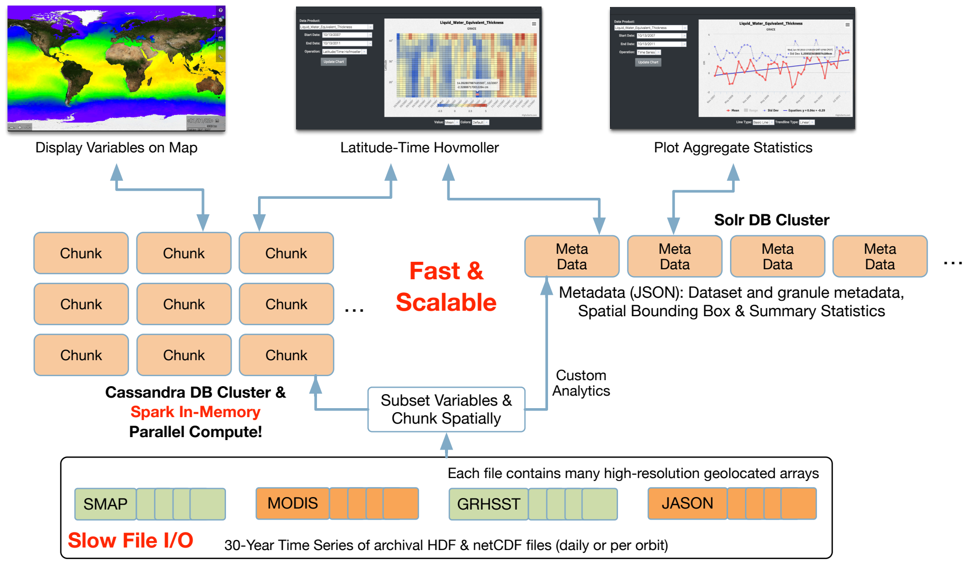
- Visualize 58 different sub-surface measurements
- Overlay with satellite measurements
- On-the-fly generation of profile measurement plot



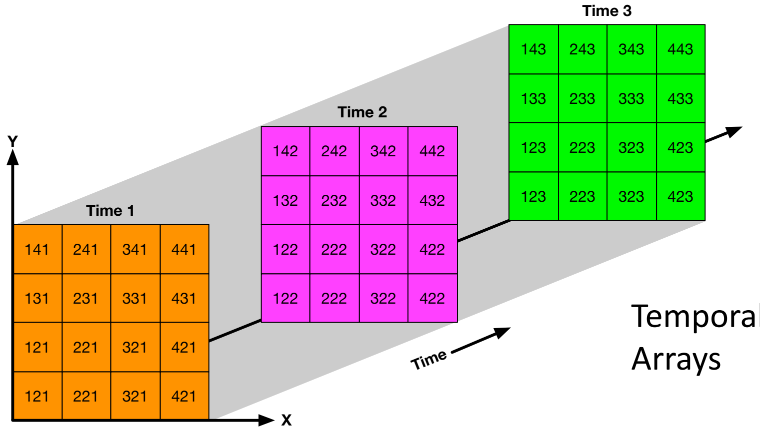


NEXUS: Scalable Data Analytic Solution

- NEXUS is a data-intensive analysis solution using a new approach for handling science data to enable large-scale data analysis
- Streaming architecture for horizontal scale data ingestion
- Scales horizontally to handle massive amount of data in parallel
- Provides high-performance geospatial and indexed search solution
- Provides tiled data storage architecture to eliminate file I/O overhead
- A growing collection of science analysis webservices using Apache Spark: parallel compute, in-memory map-reduce framework
- Pre-Chunk and Summarize Key Variables
    - Easy statistics instantly (milliseconds)
    - Harder statistics on-demand using Spark (in seconds)
    - Visualize original data (layers) on a map quickly (Cassandra store)
- **Algorithms** – Time Series | Latitude/Time Hovmöller| Longitude/Time Hovmöller| Latitude/Longitude Time Average | Area Averaged Time Series | Time Averaged Map | Climatological Map | Correlation Map | Daily Difference Average



Two-Database Architecture

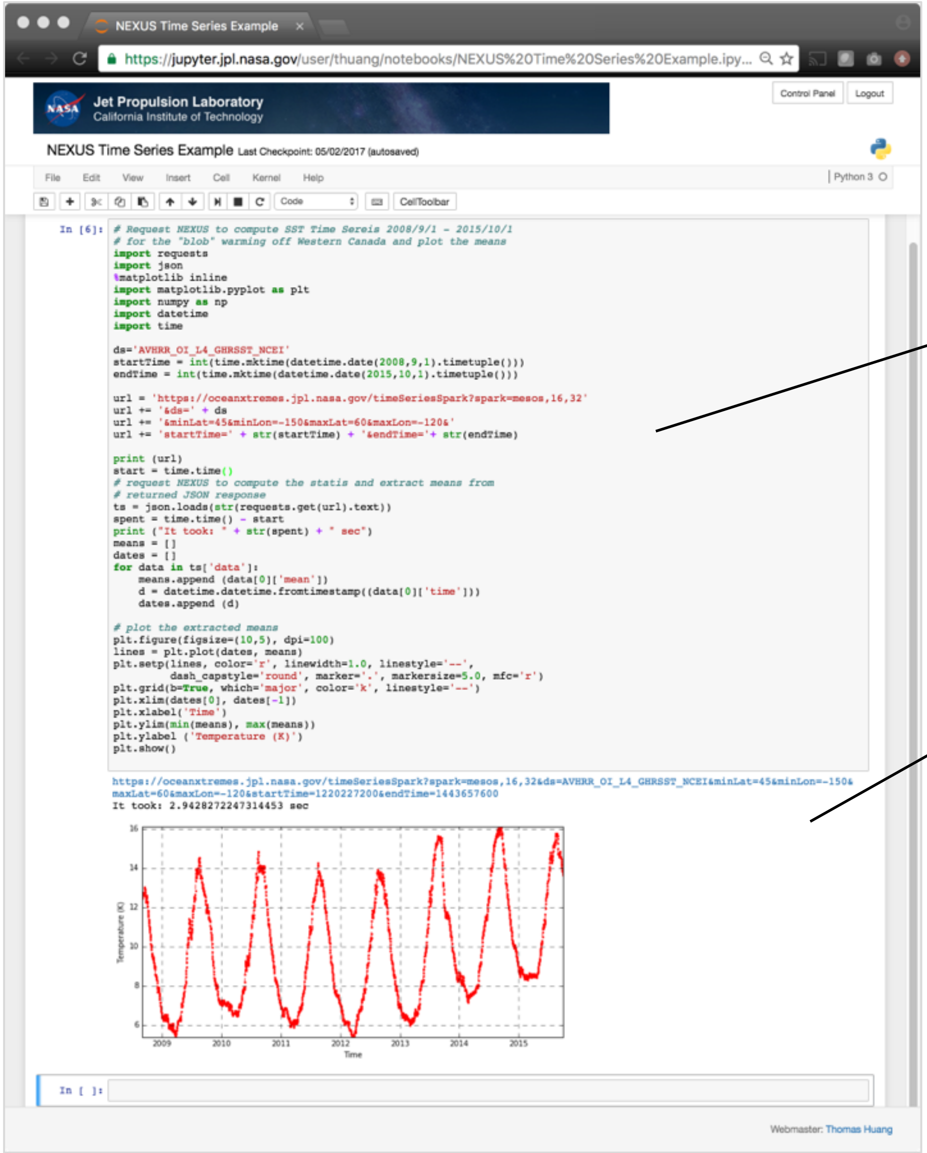


Temporal spatial Arrays

Open Source: Apache License 2
https://github.com/dataplumber/nexus

Jet Propulsion Laboratory
California Institute of Technology

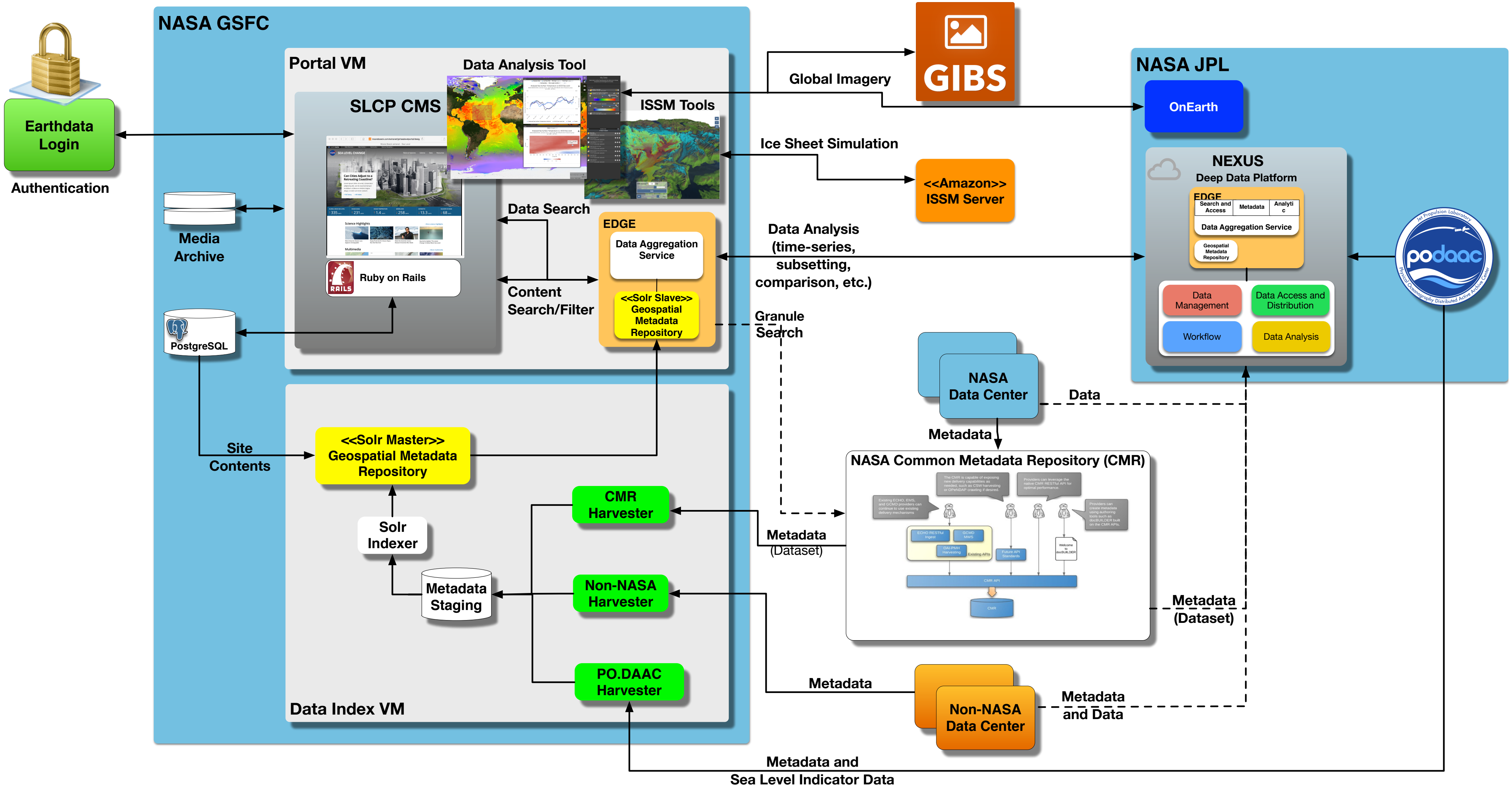
Enable Science without File Download




```
# Request NEXUS to compute SST Time Series 2008/9/1 - 2015/10/1
# for the "blob" warming off Western Canada and plot the means
…
ds='AVHRR_OI_L4_GHRSST_NCEI'

url = … # construct the webservice URL request


# make request to NEXUS using URL request
# save JSON response in local variable
ts = json.loads(str(requests.get(url).text))

# extract dates and means from the response
means = []
dates = []
for data in ts['data']:
    means.append (data[0]['mean'])
    d = datetime.datetime.fromtimestamp((data[0]['time']))
    dates.append (d)

# plot the result
…
```

```
https://oceanxtremes.jpl.nasa.gov/timeSeriesSpark?spark=mesos,16,32&ds=AVHRR_OI_L4_GHRSST_NCEI&minLat=45&minLon=-150&maxLat=60&maxLon=-120&startTime=1220227200&endTime=1443657600

It took: 2.9428272247314453 sec
```

Under The Hood



Open Source

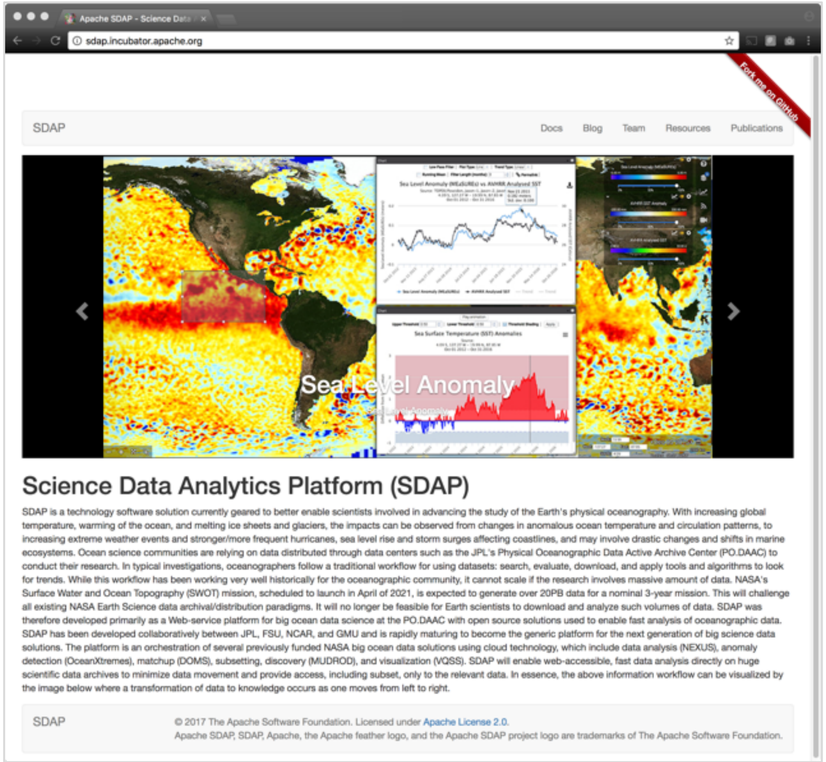
Jet Propulsion Laboratory
California Institute of Technology

Technology sharing through Free and Open Source Software (FOSS)

Further technology evolution that is restricted by projects / missions

Science Data Analytic Platform (SDAP), the implementation of **NASA AIST OceanWorks**, in **Apache Incubator**

- Cloud platform
- Analyzing satellite and model data
- In situ data analysis and coordination with satellite measurements
- Fast data subsetting
- Mining of user interactions and data to enable discovery and recommendations
- Streamline deployment through container technology



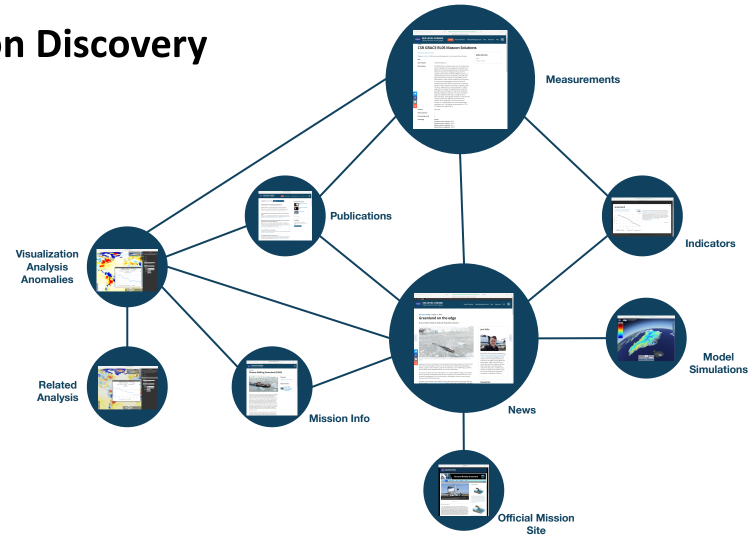
http://sdap.incubator.apache.org

Jet Propulsion Laboratory
California Institute of Technology

What to expect in 2018?

- A platform to transform data into knowledge
- Enable information discovery through dynamic ranking, recommendations, and linking
- Models and projections
- More visualizations
- Enhance support for in situ data and analysis
- Enhance support for hydrological basin analysis
- Enhance mobile experience
- More data for user to visualize and analyze
- Cloud deployment automation

Improve Information Discovery



Transforming Data to Knowledge